We Rate Dogs

Intro

WeRateDogs is an account on twitter which offers rates for cute and funny dogs.

They are known for there rates which normally exceed 10/10 and categorizing dogs as ‘doggo’, ‘floofer’, ‘pupper’ and ‘puppo’.

Insights on their tweets regarding dog ratings will be provided later, but for now, let me walk you through the process.

Gathering data

Well there are three data sets available to work with: the provided enhanced archive, the twitter API data and the neural network predictions data.

The enhanced archive data, contains data regarding the tweets like: tweet id, time, text on tweet, dog category, etc.…

The Neural Network Predictions data, is a set where the dog images had fed to a trained model to predict the object/dog breed in the image, and then returns 3 predictions and percentage of how right the prediction is

Finally, the API data, contains a lot of data like: retweet counts, favorite counts, language, content sensitivity, hashtags, mentions, etc.…
but we’re only interested in the retweet counts and the favorite counts.

Assessing and Cleaning data

It goes without question the data is not ready at all to be analyzed as it is, so after assessing the data, I’ve done some cleaning which includes:

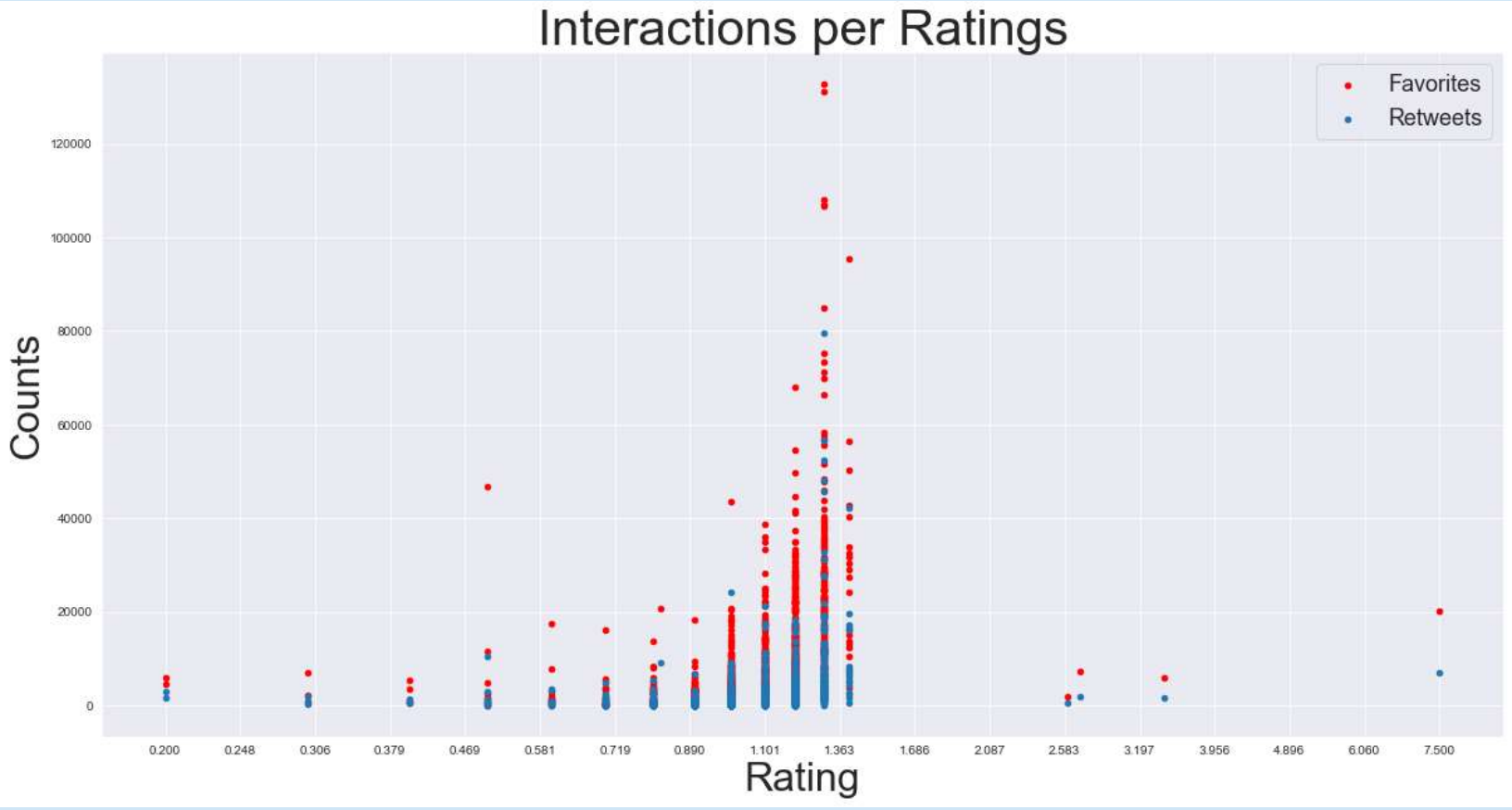
- Dropping columns with less than 2 values
- Converting timestamps from string type to datetime type
- Extracting source name and URL from html tag

- Converting ‘None’ strings to actual null values
- Excluding wrongly detected names like: a, an, the, this, etc…
- Removing data about retweets
- Removing data not concerning dogs
- Combining ratings into 1 column as a number
- Combining dog category into 1 column instead of 4
- Merging useful columns from all data sets
- Dropping columns which would be unuseful in the analysis process

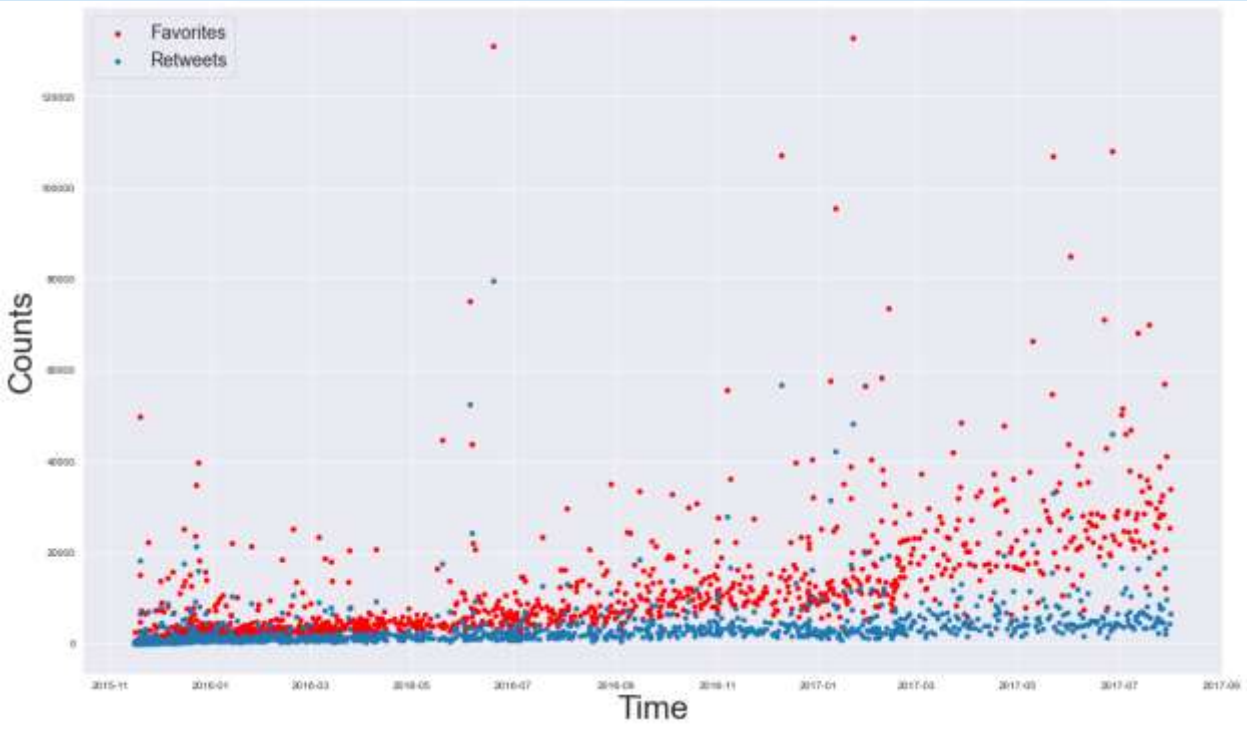
After doing the steps above, the data came out as show in the attached file.

Analyzing the data

Well, At first, what we have here is a scatter plot showing the correlation between user interactions and the ratings given by the account, it looks like the people like how most ratings are, usually around 12/10 or 13/10, but as the graph shows, when they take it too far, people to tend to like it as much, So I’d recommend staying in the normal region of ratings.

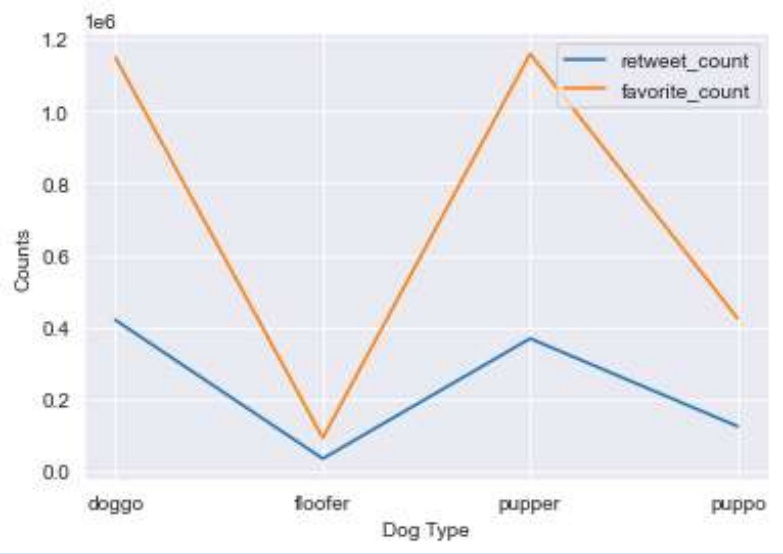


Secondly, We have the favorite counts (in red) and the retweet counts (in blue) over the span of almost 2 years, as we can see the favorites count has almost doubled in a roughly squared correlation with time, but on the other hand, the retweets count is almost the same or with a very small increase, which tells us that the tweets need to be more engaging with the users to increase the retweets count.

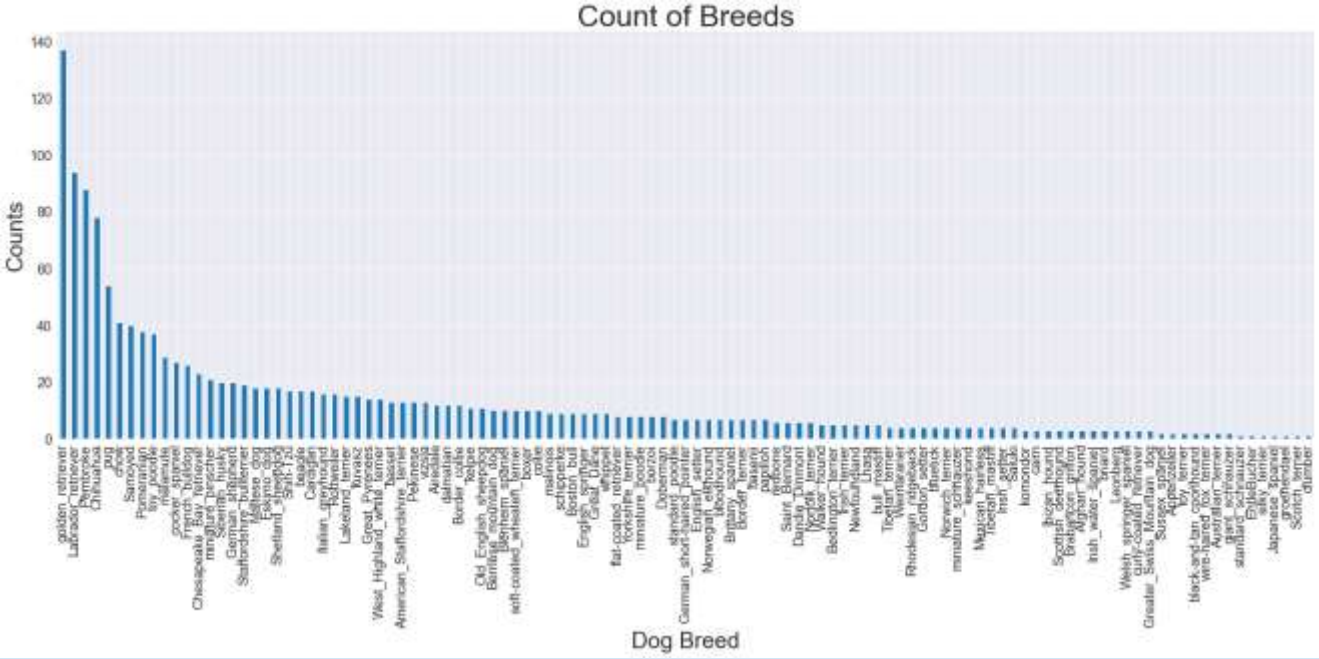


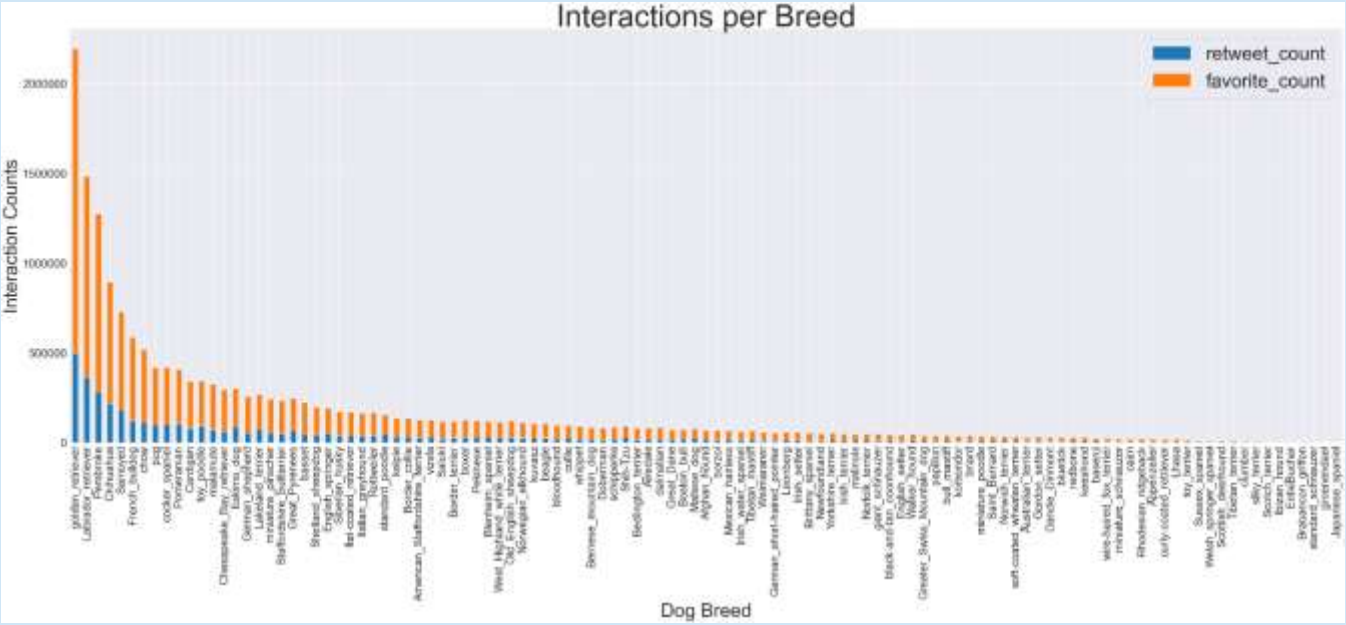
On the right here, is a simple line plot which tells us that people are much more likely to interact with the 'doggo' and 'pupper' category of dogs than with others.

Whatever the reason for this is, increasing posts about those 2 categories would definitely increase the user interactions.



Finally, the two following bar plots, show the count of breeds and the interactions per breed, respectively.





Both graphs show that, the Golden Retriever is both the most popular and has the most counts, with the Labrador Retriever in second place and the Pembroke Welsh Corgi in third, so increasing tweets about these breeds would most likely invite more user interaction.